VU VRIJE UNIVERSITEIT AMSTERDAM

VU Research Portal

Effective repairs in survey interviews

van der Zouwen, J.; Smit, J.H.

published in
Quality and Quantity
2006

DOI (link to publisher)
10.1007/s11135-005-6102-2

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
van der Zouwen, J., & Smit, J. H. (2006). Effective repairs in survey interviews. *Quality and Quantity*, *40*(2), 245-258. https://doi.org/10.1007/s11135-005-6102-2

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 04. May. 2023

Quality & Quantity (2006) 40:245–258
DOI 10.1007/s11135-005-6102-2
© Springer 2006

Effective Repair in Survey Interviews

JOHANNES VAN DER ZOUWEN[1,⋆] and JOHANNES H. SMIT[1,2]
[1]*Faculty of Social Sciences;* [2]*Faculty of Medicine, Vrije Universiteit Amsterdam*

Abstract. In standardized survey interviews with closed-ended questions it quite often occurs that respondents do something else than mentioning their choice from the set of response categories presented together with the question. In these cases their initial response is inadequate and interviewers have to initiate repair activities.
By analyzing 804 question–answer sequences, stemming from parts of a survey with elderly people ($N = 201$), we found that 322 (41%) of the initial responses were inadequate. In 74% of these 332 sequences the interviewers made an attempt at repair. Of all repair attempts, 75% were performed correctly, that is, in a nondirective way. Most of the repair attempts (86%) result in an adequate final response of the respondent. The result of the repair activities by the interviewers is that the number of sequences with an adequate response increased from 472 to 688, thus an increase of about 45%.
There does not exist 'one best way' of repairing. The most important thing is that *some* repair initiative has to be taken by the interviewer and that the repair is done in a nondirective way.

Key words: interviewer behavior, closed question, adequacy of response, mismatch answers, repair, probing

1. Introduction

In spite of the increasingly high costs of interviewer administered questionnaires and the rapid development of self-administered (electronic) alternatives, telephone and face-to-face interviews are sometimes the only, or at least the best, solution for the researcher to collect data about complex phenomena, or from difficult to reach subjects.

An efficient way to collect relevant and complete information in surveys is the use of standardized closed questions in the questionnaire. Here not only the wording of the question is standardized, but also the set of response alternatives to be presented by the interviewer.

Ideally, the respondent only mentions his[1] choice of the set of presented response alternatives, whereupon the interviewer can proceed to the

⋆ Prof Johannes van der Zouwen, Department of Social Research Methodology, Faculty of Social Sciences, Vrije Universiteit, De Boelelaan 1081, 1081 HV, The Netherlands. E-mail: j.van.der.zouwen@fsw.vu.nl

next question. However, in methodological research it was found that this ideal is not always reached in practice (G. Loosveldt, 1985, Unpublished PhD dissertation; Pruefer and Rexroth, 1985; Sykes and Morton-Williams, 1987; Sykes and Collins, 1992; Van der Zouwen and Dijkstra, 1996; Houtkoop-Steenstra, 2000; Van der Zouwen and Smit, 2004).

Deviations from this ideal can occur because the respondent does not understand (parts of) the question or of the response options presented, and asks the interviewer for an explanation. Or the question and the response options do not fit the situation of this particular respondent. In these cases the respondent has to 'negotiate' with the interviewer how to solve this problem. Moreover, in cases that a question is followed by an instruction, it may be unclear to the respondent when he is expected to start answering, thus leading to interruptions and overlapping speech. Finally, it is also possible that the question itself induces responses that are completely normal within the context of a conversation, but that are nevertheless not equal to one of the response options presented, and therewith inadequate.

This means that quite often respondents do something else than mentioning their choice from the set of response alternatives; so their initial response is inadequate. In these cases the interviewer is instructed to intervene and try to obtain, by further probing, as yet a proper response from the respondent. In the words of Moore and Maynard (2002: 300): "..whenever a respondent's answer fails to reproduce one of the prespecified answer options, interviewers should probe it, that is, initiate repair and invite the respondent to redo the answer by selecting a single option."

This probing task is not only necessary, it is also difficult to perform, and incorrect probing by the interviewers may have negative effects on data quality. Or, as Kahn and Cannell (1957: 224) have put it nearly half a century ago, "..it is clear that the inadequate response makes severe demands on the interviewer's skill and judgement. Errors and omissions in the technique of probing must be counted among the major causes of interviewer bias."

Probing in order to repair an inadequate initial response may be a difficult and essential part of the task of the interviewer, yet little is known about how this task is performed, and what the effects are. "What actually happens during probing has received little study. These negotiations, however, have considerable impact on the social atmosphere of the interview and the reliability and validity of the resulting data" (Schaeffer, 1991: 388).

The present paper aims at getting insight into the different types of deviations by respondents from the ideal of the standardized survey interview, the frequency by which these types occur, the ways in which interviewers attempt at repair, and the strategies that appear to be effective.

2. Types of Inadequate Initial Responses

The researcher, who includes a standardized closed question in a questionnaire for a telephone or face to face survey, does so on the basis of a set of (mostly implicit) assumptions. These assumptions are: (1) the respondent is *willing* to answer the question, and (2) he has the necessary *information* to do so. Moreover, (3) his reaction is an answer to precisely this *particular* question, and not to another one, (4) he answers the question by mentioning one of the presented *response alternatives*, and (5) he chooses the response alternative that *best fits* with his particular state. Finally, (6) the interviewer *correctly codes* the answer given by the respondent. If all these assumptions appear to be valid, the response code, or the score, contains correct and relevant information regarding the value this respondent has on the variable the researcher wants to measure with this question.

However, it turns out that during survey interviews departures from these assumptions occur. For example, the respondent *refuses* to answer the question; he reacts with *I don't know*; or gives an answer that is, at least in relation to the question posed, *irrelevant*. A fourth departure is that a relevant answer is given, but *not* in the form of a choice from one of the presented response alternatives.

If interviews are recorded on tape (and preferably transcribed into interview protocols) it becomes possible to investigate to what extent the answer that the respondent (initially) gave to the question from the questionnaire, forms a departure from one of these assumptions.

The initial answer of the respondent, given to a particular question from the questionnaire, will be assigned one of the following codes:

- RF (refusal): the respondent refuses to answer the question;
- DK (don't know): the respondent says that he does not know the answer;
- IA (irrelevant answer): the response is not (fully) related to the question posed;
- MM (mismatch answer): the (relevant) answer of the respondent is not equal to one of the response alternatives provided;
- If the initial answer of the respondent seems to fulfil all the requirements of an adequate answer, it is coded as AA (adequate answer).

3. The Data

The data used in this paper were collected as part of the Longitudinal Aging Study Amsterdam (LASA), a study of antecedents and consequences of changes in autonomy and well-being among older adults (see Beekman et al., 1995). In 1992 a representative sample of 3805 adults aged 55 to 85, stratified for age and gender, was drawn from the population registers of

11 municipalities in the Netherlands. The data for the present study were collected by CAPI in 1998. At that time 1770 participants were still in the study and capable of completing a face-to-face interview of approximately 2 hours.

Interviewers were recruited from the area the respondents lived in. The training included reading assignments on interviewer techniques, videotape exercises, pilot testing of instruments, and a complete interview with an elderly person. The interviewers were mostly female (95%), between 25 and 60 years old, and had a medium level of education (10 years).

All interviews were audiotaped for review by the fieldwork staff in order to give supportive feedback on interviewer behavior (Van der Zouwen and Smit, 2004). We will use the transcriptions of the question-answer sequences following four closed questions from the LASA-questionnaire, stemming from a random set of 201 interviews (see Box 1).

Box 1. The four selected questions

[If the respondent has no partner]
Q3a: Could you indicate, by means of this card, in which category your NET income falls? Please mention the number of the category that applies.
[If the respondent has a partner]
Q3b: Could you indicate, by means of this card, in which category the joint NET income of you and your partner falls? Please mention the number of the category that applies.
N.B. The card contains 12 income categories (excluding vacation allowance) ranging from 1000 to 1250 guilders per month (12,000–15,000 guilders per year) to 5001 guilders per month or more (60001 guilders per year or more).
[If the respondent has no partner]
Q4a: Did your income decrease in the last three years?
[If the respondent has a partner]
Q4b: Did your income decrease in the last three years? Do NOT count the income of your partner.
0, don't know; 1, no; 2, yes, less than 100 guilders net per month; 3, yes, 100–200 guilders net per month; 4, yes, 200–300 guilders net per month; 5, yes, 300–400 guilders net per month; 6, yes, 400–500 guilders net per month; 7, yes, more than 500 guilders net per month.
[If the respondent has no partner]
Q6a: Are you satisfied with your income level?
[If the respondent has a partner]
Q6b: Are you satisfied with your income level? We mean the joint income of you and your partner. 1, dissatisfied; 2, a little dissatisfied; 3, neither satisfied nor dissatisfied; 4, a little satisfied; 5, satisfied.
Q8: Do you expect that in the next 2 years your income will 1, decrease?; 2, remain the same?; 3, increase?

4. Different Questions, Different Deviations

The four survey questions from Box 1 were selected because we knew from an earlier study (Van der Zouwen and Smit, 2004) that they generate a considerable number of inadequate initial responses. Moreover, these four questions together represent all types of inadequate initial responses.

Table I shows, for each of the four selected questions, the distribution of the sequences over the adequate initial answers (AA) and over the four types of inadequate responses (RF, DK, IA, and MM).

Of the 804 sequences analyzed, the majority (59%) receives code AA, because the interviewer got an adequate initial answer from the respondent. However, in 41% of the sequences the initial answer is inadequate. Of the 332 sequences with such an inadequate initial answer, nearly 66% is a mismatch answer (MM) and 25% is a 'don't know' answer (DK). The remaining 9% are divided over refusals (RF) and irrelevant answers (IA).

Compared with the data from Brenner (1982) and from Sykes and Morton-Williams (1987), the proportion of inadequate initial responses to these four questions is quite high. This may be partly due to the fact that the respondents were elderly people (Bradburn and Sudman, 1979: 49), and that the questions were sometimes hard to answer, due to a difficult topic or ill-designed question wording. Our finding that mismatch answers (code MM) form the majority of the inadequate initial responses, is in agreement with the results of the study by Ongena (2001) and Dijkstra and Ongena (2002).

From Table I it appears that the distribution over the types of inadequate initial answers is quite different for each of the four survey questions[2] ($\chi^2 = 224.1$; df = 12; $p < 0.001$). Question Q3 [about the net income of the respondent (and his or her partner)] is a (privacy) sensitive question that is often hard to answer. This question thus leads to relatively many refusals (RF = 3%) and don't know responses (DK = 19%).

Table I. Distribution (%) of the sequences over adequate initial answers (AA) and over four types of inadequate initial answers, by question

Initial answer	Q3	Q4	Q6	Q8	Total
Adequate initial answer (AA)	51	81	38	65	59
Refusal (RF)	3	0	0	0	1
Don't know (DK)	19	5	1	17	11
Irrelevant answer (IA)	1	3	2	7	3
Mismatch answer (MA)	27	11	59	11	27
Total (%)	100	100	100	100	100
N	201	201	201	201	804

Question Q4 ("Did your income decrease in the last three years?") looks like a simple yes/no-question. But the set of response alternatives is much larger, because the 'yes' answers are further subdivided according to the amount of income decrease. This question generates relatively many irrelevant initial answers (IA = 3%).

Question Q6 ("Are you satisfied with your income level?") is a common satisfaction question. The problem here is that this yes/no-question is followed by five response alternatives. This means that quite often the initial response 'yes', or 'no', has to be repaired into one of the five response categories. This question thus generated many mismatches (MM = 59%).

Question Q8 ("Do you expect that in the next two years your income will decrease, remain the same, or increase?") is a hypothetical question, relating to the future, and as such hard to answer. Moreover, for quite a few respondents the term 'income' relates to their 'buying power'; and they motivate their choice of a response category accordingly. This question thus generated relatively many don't know responses (DK = 17%) and irrelevant answers (7%).

We can conclude that the distribution over the different types of initial answers is very uneven, and this distribution is strongly related to the topic and the format of the selected survey questions.

5. Attempts at Repair by the Interviewer

As mentioned above, 332 of the 804 selected sequences have a code for an inadequate initial answer. The protocols of these over three hundred sequences will be used for the further analysis.

Because inadequate responses may hamper the completeness and correctness of the data collected, interviewers are usually instructed by the researcher to try to change the inadequate response of the respondent into an adequate one, in other words, to probe in an attempt to repair the initial response.

From Table II it appears that, on average, in about three-quarter (74%) of the sequences in need for repair, the interviewer indeed does an attempt at repair. In the few cases in which the initial response was a refusal (code RF), interviewers made relatively less attempts at repair. The differences with respect to the percentage of attempts at repair between the four different types of inadequate initial answers are not significant ($\chi^2 = 5.1$; df = 3; $p > 0.05$).

6. The Correctness and Effectiveness of Attempts at Repair

It is quite difficult to decide whether an attempt at repair by the interviewer may be called 'correct' or 'incorrect.' Viterna and Maynard (2002)

Table II. Types of inadequate initial answers and absence or presence of an attempt at repair by the interviewer (in %)

	Attempt at repair			
Initial answer	No	Yes	Total (%)	*N*
Refusal (RF)	60	40	100	5
Don't know answer (DK)	19	81	100	84
Irrelevant answer (IA)	28	72	100	25
Mismatch answer (MM)	28	72	100	218
Total	26	74	100	332

have shown that survey research centers differ considerably with respect to the probing strategies that are instructed and trained in these centers. However, all survey researchers agree that the interviewer has to *probe* inadequate answers, and that this probing has to be done in a *nondirective*, or neutral, way, in order not to affect the final response of the respondent. But survey methodologists disagree with respect to which probing strategies can be classified as directive or nondirective.

Fowler and Mangione (1990: 39–40) – in line with Brenner (1982) – are very strict as regards the strategy for repairing a mismatch answer: the interviewer is not allowed to code the answer into a response category. And "the other mistake an interviewer can make in probing a closed-ended question is not to repeat all the alternatives when the alternatives need to be repeated" (..). "A truncated version of the set of responses is not the same stimulus, and it will affect the answers." Other survey researchers are less strict, and think that repeating only a well-chosen subset of response categories, based on the initial answer of the respondent, together with a probe question like "which one comes closest to your opinion?" is not necessarily directive.

Moore (2000: 13) discusses a probe question that is quite often used by interviewers after listening to the initial answer of the respondent: "Is it X?" or "Would you say X?". Such a

> "*verification probe* is somewhere between the withholding of probing and nondirective probing. It requires the respondent to ratify the answer-option before it gets recorded, and at the same time, acknowledges that the respondent already provided a conversationally adequate answer. Although this verification probe technically violates standardization procedures (Fowler & Mangione, 1990), it nonetheless appears to preserve their intent. Furthermore, there may be good interactional reasons for violating standardization procedures in

these cases. Correcting the form of an otherwise understandable and adequate utterance can come off as "annoying" or even "rude"."

Because empirical evidence is lacking for making the best choice between different probing strategies, we adopted the 'practical' argumentation of Moore and therefore take the quite 'liberal' view that a verification probe can be counted as a correct repair strategy.

An attempt at repair is called 'successful' or 'effective' if the initial inadequate answer of the respondent is changed into an adequate one. For example, a question initially answered with "I don't know" receives as yet a substantive response. Or the respondent eventually chooses one of the prespecified response alternatives. It is possible that an attempt at repair, correct or not, does not succeed. For example, the respondent sticks to his refusal to answer the question. On the other hand, it is also possible that an inadequate initial response is corrected by the respondent himself ("Oh now I remember, it was X"), possibly without any intervention by the interviewer. This means that missing repair still may be followed by an adequate final response, whereas attempts at repair may remain unsuccessful, i.e., are followed by an inadequate final response.

From Table III it appears that of all repair attempts, 75% are performed correctly. This percentage is lower for question Q4 (61%), and higher for question Q3 (81%). Most of the repair attempts, 86%, result in an adequate final response of the respondent. The repair attempts concerning question Q6 are the most successful: 93%.

There exists a strong relationship between the correctness of the attempts at repair and their effectiveness: correct attempts at repair have a significantly higher probability of being successful than incorrect ones: Gamma = 0.48 ($p < 0.001$).

The information contained in Tables I through III can be summarized in the following way: Of the 804 sequences that form the basis of our analysis, 472 receive code AA, because the respondent directly gives an

Table III. Percentage of correct and incorrect, and successful and unsuccessful, attempts at repair for each of the four questions

Repair attempt	Q3	Q4	Q6	Q8	Total
Correct and successful	70	45	72	70	67
Correct and unsuccessful	11	16	1	6	8
Incorrect and successful	17	26	21	14	19
Incorrect and unsuccessful	2	13	5	10	6
Total	100	100	100	100	100
N	89	31	76	50	246

adequate answer to the question from the questionnaire. The remaining 332 sequences receive one of the codes for inadequate initial response (RF, DK, IA or MM).

Of the 332 sequences that are in need of repair, 246 indeed contain an attempt of the interviewer at repair. In these 246 sequences, the repair attempt is performed correctly in 185 cases, and incorrectly in 61 cases. Of the 185 sequences with a correctly performed attempt at repair, 166 result in an adequate final response, while 19 remain inadequate. Of the 61 sequences with an incorrect attempt at repair, 46 eventually lead to an adequate final response, and 15 remain inadequate. Of the 86 sequences that are in need of repair, but do miss an attempt at repair of the interviewer, four eventually lead to an adequate final response, while 82 remain inadequate. With these data Table IV can be constructed.

Table IV shows that there exists a very strong relationship ($Q = 0.98$; $p < 0.001$) between making an attempt at repair and getting an adequate final response: repair does help indeed. And 'spontaneous' repair, without an intervention by the interviewer, hardly occurs.

The result of the repair activities of the interviewer is that the number of sequences with an adequate response is increased from 472 to 784, an increase of about 46%. However, it must be noted that part of this increase is the result of incorrect repair, especially by using leading probes.

7. Types of Effective and Correct Repair Activities

The 166 sequences that belong to the first row of Table III need further analysis because these sequences ended up with an adequate final response after a correct attempt at repair by the interviewer. An analysis of these sequences may show us what kinds of correct repair strategies appear to be effective.

We assume that the strategy used by the interviewer to repair an inadequate initial response, and its effectiveness, is strongly related to the type of inadequate initial response that has to be repaired. And we already observed (in Section 4) that the distribution over the types of inadequate

Table IV. The effectiveness of attempts at repair

	Attempt at Repair		
Result	Yes	No	Total
Adequate final response	212	4	216
Inadequate final response	34	82	116
Total	246	86	332

initial responses is strongly related with the topic and format of questions from the questionnaire. We have therefore performed an analysis of the sequences belonging to the first row of Table III (the correctly and successfully repaired sequences) for each cell of this row, that is, for each of the four questions, separately. The results of this analysis are summarized in Box 2.

For each question the problems for the respondent (abbreviated with R) are mentioned, followed by the type of correct and successful repair strategies that were frequently used by the interviewers (abbreviated with I). The most frequently used repair strategies are printed bold face.

8. Differences Between the Interviewers

In earlier studies (Van der Zouwen and Smit, 2003, 2004), it appeared that quite a lot of inadequate responses were initiated by inadequate wording of the question or by a misfit between the question and its response alternatives. For example, the many mismatch answers (code MM) following question Q6 are caused by the fact that a yes/no question ("Are you satisfied with your income level?") is followed by five response alternatives, making the logical 'yes' or 'no' answer inadequate. By redesigning the questionnaire, many inadequate initial responses, and therewith repair activities, can be prevented.

However, part of the inadequate initial responses, cannot be prevented by improvement of the questionnaire, because they are related to the complexity of the topic of the question, or to the restricted cognitive abilities of the respondent. In these cases repair remains necessary. And here the competence of the interviewer may play an important role.

To investigate differences between interviewers concerning their competence, we selected from the set of 332 sequences with an inadequate initial response, those 156 that were conducted by the nine interviewers of whom eight or more interviews belonged to this set.[3]

The percentage of sequences with an inadequate initial response did not differ significantly ($p > 0.05$) between interviewers. However, the percentage of sequences in need of repair, for which the interviewer actually made an attempt at repair varied from 50% to 94%. These latter differences appear to be significant: $\chi^2 = 19.3$; df = 8; $p = 0.013$. Further analysis showed that these differences are mainly caused by differences in interviewer behavior that show up when question Q6 ("Are you satisfied with your income level?") is answered with 'yes'. Some interviewers systematically probe further by mentioning all response alternatives, or a selection thereof, while others take the response 'yes' as meaning the response alternative 'satisfied'. Excluding sequences related to question Q6 from the analysis,

Box 2. The main problems the four questions pose for the respondents (R) and correct and effective repair strategies used by the interviewers (I)

Problems for R	Repair strategies used by I
Q3 "Could you indicate, by means of this card, in which category your NET income falls?"	I uses documents and information from the partner of R (3)
Income is a sensitive topic	**I stresses that an 'estimation' is sufficient (7)**
R often lacks information about income (of partner)	**I helps R by making the calculation of the total net income (12)**
Difficult to answer question: various sources of income; the combination of monthly payments with annual payments; including income of partner; transforming gross income into net income; amounts to be translated into the response codes on the show card	I repeats question or answer (3) I encourages R to make the calculations (4) I mentions a subset of options based on the initial response (2) **I makes a verification probe (17)** **I 'translates' amount of income into the corresponding response code (11)**
Q4 "Did your income decrease in the last 3 years? Do NOT count the income of your partner"	I uses information from partner (1)
Question is difficult to answer for R with a partner, whose own income has been decreased	**I eliminates misunderstandings (4)** I makes a verification probe (1) I encourages R to answer the question (3)
Small income categories	I repeats the question (1)
Concept "decrease of income" is unclear to R	I 'translates' the amount of decrease of income into a response code (1)
No use is made of a show card	Other strategies used by I (3)
Q6 "Are you satisfied with your income level?" (dissatisfied; a little dissatisfied, neither satisfied nor dissatisfied, ...)	I repeats the question (1) I explains the question (1)
Response options not included in text of the question	**I mentions as yet the five response options (52)** I corrects R (1) I makes a verification probe (1)
Q8 "Do you expect that in the next two years your income will decrease, remain the same or increase?"	I repeats the question (2)
Question is difficult to answer: what is 'income'?; and it is the government that makes the decisions about the pensions	**I probes further: "but what do you expect?" (26)** I makes a verification probe (1) I explains the question (3) I corrects R (3)

makes that the differences between the interviewers with regard to repair activities, become insignificant.

The differences between the interviewers are related to their judgment whether a repair activity is needed or not. But as soon as they start a repair activity, they do not differ significantly with respect to the correctness by which the repair attempt is performed. Thus the difference between the interviewers lies in their judgment about the *need* for repair, not in the *quality* of the repair.

9. Summary and Conclusion

In standardized survey interviews with closed-ended questions it quite often occurs that respondents do something else than mentioning their choice from the set of response categories presented together with the question. In these cases their initial response is inadequate and interviewers have to initiate repair activities.

The repair task of interviewers is both necessary, and difficult to perform. There exists hardly any literature about how interviewers actually perform this task. The present paper was intended as a modest contribution to fill this lack of information.

By analyzing 804 question-answer sequences, stemming from parts of survey interviews with elderly people, we found that 41% of the initial responses was inadequate. The distribution over the types of inadequacies observed (refusal, don't know answer, irrelevant answer or mismatch answer) differed strongly for the four selected questions, depending on their topic and format.

In 74% of the 332 sequences in which the initial response is inadequate, the interviewer made an attempt at repair. Of all repair attempts, 75% were performed correctly, that is, in a nondirective way. Most of the repair attempts (86%) result in an adequate final response of the respondent. The result of the repair activities by the interviewers is that the number of sequences with an adequate response increased from 472 to 684, thus an increase of about 45%.

A detailed analysis of those sequences that were correctly and successfully repaired showed large differences between the four questions. An explanation for these differences can be found in the fact that the problems the different questions posed to the respondent vary a lot and therewith the type of inadequate response that has to be repaired. There does not exist 'one best way' of repairing. The most important thing is that *some* repair initiative has to be taken by the interviewer and that the repair is done in a nondirective way.

Acknowledgements

This research is based on data collected in the context of the Longitudinal Aging Study Amsterdam (LASA). This program is conducted at the Vrije Universiteit in Amsterdam and is supported by the Ministry of Health, Welfare, and Sports.

Notes

1. For the sake of readability we will indicate the respondent and the researcher with 'he', and 'his,' and the interviewer with 'she' and 'her'.
2. The sequences we have analyzed are not fully independent of each other. Some belong to the same interview, and some of the interviews are conducted by the same interviewer. A multi-level analysis is not possible with our data, because of the small number of cases. We treat the sequences as if they are independent but use a quite conservative level of significance ($p < 0.01$).
3. To guarantee the reliability of the results, we excluded from the analysis those interviews that were conducted by interviewers who did only a few interviews.

References

Beekman, A. T., Deeg, D. J. H., van Tilburg, T., Smit, J. H., Hooijer, C. & Van Tilburg, W. (1995). Major and minor depression in later life: A study of prevalence and risk factors. *Journal of Affective Disorders* 36: 65–75.

Bradburn, N. M., Sudman, S. & Associates (1979). *Improving Interview Method and Questionnaire Design*. San Francisco: Jossey-Bass.

Brenner, M. (1982). Response effects of 'role-restricted' characteristics of the interviewer. In: W. Dijkstra & J. van der Zouwen (eds.), *Response Behaviour in the Survey-Interview*. London: Academic Press.

Dijkstra, W. & Ongena, Y. (2002) Evaluating Questionnaires by Analysing Question–Answer Sequences. Paper presented at the International Conference on Questionnaire Development, Evaluation and Testing Methods, November, 14–17, 2002, Charleston SC.

Fowler, Jr., F. J. & Mangione, T. W. (1990). *Standardized Survey Interviewing; Minimizing Interviewer-related Error*. Newbury Park: Sage.

Houtkoop-Steenstra, H. (2000). *Interaction and the Standardized Survey Interview; The Living Questionnaire*. Cambridge: Cambridge University Press.

Kahn, R. L. & Cannell, C. F. (1957). *The Dynamics of Interviewing: Theory, Tactics and Cases*. New York: Wiley.

Moore, R. J. (2000). Differences in probing practice at two U. S. survey centers. In: J. Blasius, J. Hox, E. de Leeuw & P. Schmidt (eds.), *Social Science Methodology in the New Millenium*, Proceedings of the Fifth International Conference on Logic and Methodology, held at Cologne, Germany, October 3–6, 2000. Opladen, Germany: Leske & Budrich, CD-ROM.

Moore, R. J. & Maynard, D. W. (2002). Achieving understanding in the standardized survey interview: Repair sequences. In: D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer & J. van der Zouwen (eds.), *Standardization and Tacit Knowledge; Interaction and Practice in the Survey Interview*. New York: Wiley.

Ongena, Y. P. (2001), *Problematic Deviations in Question–Answer Sequences*. Amsterdam: Vrije Universiteit; Internal report.

Pruefer, J. & Rexroth, M. (1985). Zur Anwendung der Interaction-Coding Technik. *ZUMA-Nachrichten* 17: 2–49.

Schaeffer, N. C. (1991). Conversation with a purpose – or conversation? Interaction in the standardized interview. In: P. P. Biemer, R. M. Groves, L. L. Lyberg, N. A. Mathiowetz and S. Sudman (eds.), *Measurement Errors in Surveys*. New York: Wiley.

Sykes, W. & Morton-Williams, J. (1987). Evaluating survey questions. *Journal of Official Statistics* 3: 191–207.

Sykes, W. & Collins, M. (1992). Anatomy of the survey interview, *Journal of Official Statistics* 8: 277–291.

Van der Zouwen, J. & Dijkstra, W. (1996). Trivial and non-trivial question–answer sequences; Types, determinants and effects on data quality. In: *Proceedings of the International Conference on Survey Measurement and Process Quality in Bristol* (UK), Alexandria, VA: American Statistical Association.

Van der Zouwen, J. & Smit, J. H. (2003). Control processes in survey interviews: A cybernetic approach. Paper presented at the Fourth International Conference on Sociocybernetics, June 29 – July 5, 2003, Corfu, Greece.

Van der Zouwen, J. & Smit, J. H. (2004). The diagnostic approach: Evaluating survey questions by analyzing patterns of behavior codes and transcripts of question-answer sequences. In: S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin and E. Singer (eds.), *Methods for Testing and Evaluating Survey Questionnaires*. New York:Wiley.

Viterna, J. S. & Maynard, D.W (2002). How Uniform is Standardization? Variation Within and Across Survey Research Centers. In: D. W. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer and J. van der Zouwen (eds.), *Standardization and Tacit Knowledge; Interaction and Practice in the Survey Interview*. New York: Wiley.